

Identifying Factors in Congressional Bill Success

CS224w Final Report

Travis Gingerich, Montana Scher, Neeral Dodhia

Introduction

During an era of government where Congress has been criticized repeatedly as being the most polarized and unproductive in American history [7, 8], understanding group decision making and the power dynamics behind it has peaked our interest. Thus our project investigates the factors that are involved in the success of Congressional bills, and how we can use this data to develop a model for predicting the success of a bill.

Several metrics and techniques can be used to investigate the factors that lead to a bill's success. We start by analyzing some of the basic and inherent properties of bills, i.e. static factors that could be calculated upon the introduction of a bill. This includes: date bill is introduced, how partisan its co-sponsors are, the relative rank of its sponsor, and other features. We also investigate network centrality in collaboration networks between Congress members, using both committee membership and rank in committee to generate these graphs. We used both PageRank and betweenness centrality to try and understand a legislator's influence within this network. The last static property we explored was the connection between bill success and campaign finance sources for which we developed a tripartite graph linking campaign finance sources to candidates to bills they sponsored.

We also use the congressional voting data as another, dynamic component in predicting bill success. We are interested in modeling individual voting behavior of Congress members and investigating how Congress members relate to a particular bill across several dimensions. To do this, we apply matrix factorization techniques, widely used in recommender systems, to model characteristics of both proposed bills and legislators.

In order to establish a basis for prediction we use a number of the aforementioned properties, based on their strength as bill outcome indicators, as features in a machine learning model. We validate this model by splitting historical data on bills as well as voting records available from GovTrack.us into a training set and a test set.

Prior Work

GovTrack.us [1] has performed similar work in analysing factors that contribute to a bill's success. With around fifty features, many of which are hyper specific, e.g. specific phrases mentioned in a bill, they trained a logistic regression classifier to come up with a bill prognosis. The most interesting factor they use is a *leadership* score which is defined by the legislator's PageRank in a bill co-sponsorship network. As this paper used a similar dataset it is very relevant to our work, however we focus more on the congressional network structure and relationship among congress members and less on the restrictive textual attributes of a given bill.

In [2], Leskovec, Huttenlocher, and Kleinberg study the Wikipedia promotion process as a model for group decision making through online social media. Leskovec et al. analyze forms of relative assessment to understand how individual users will vote on a candidate and how those votes will aggregate to form a final election decision. They use several "figures of merit," to show that there are non-monotonic effects of relative merit and a clear difference in voting patterns when a voter has less merit versus more merit than the candidate. This relates to our goal of developing objective measures of each congressman, how the congressmen relate to each other, and how that relationship influences both a congressman's vote or a bill they sponsor.

Guha, et al. in [3], investigate the flow of trust (and distrust) through the web. On a dataset from the Epinions website, they modeled the spread of trust as a matrix operation. Their factors included how trust is propagated: by transitivity (direct propagation), by symmetry (transpose), by implication of shared beliefs (co-citation), etc; whether distrust is propagated and, if so, how. There were several factors, particularly homophily, that the paper

omitted from its investigations. This project attempts to capture this as the dimensions of a bill (their sponsor, date of introduction, etc).

Paterek and Arkadiusz in [6] describe several techniques used in the Netflix Prize competition, which uses historic ratings data to provide better recommendations. Paterek et al present several models including: regularized singular value decomposition (also referred to as matrix factorization in other works) of a user-rating matrix, K-means using vectors of user ratings, and using K-nearest neighbors with the results of matrix factorization. The most successful techniques built upon matrix factorization. The main weakness in their methods was limited information about the user or movies. For example, including data on demographics, users’ connections, genre, age and running time could improve the predictions. This paper was most relevant in providing a concrete example of an application of matrix factorization and allowing us to use their method as a baseline while adapting the techniques described to suit the context.

Dataset

The majority of our data including the voting record of each Congress member, committee membership data, detailed information about each legislator and detailed information on bills, including their current status, past actions, sponsors and summary of votes, comes from GovTrack.us. Data on financial contributions is available from the Federal Electoral Commission's website.

While there are several types of bills, this project looks only at house, senate, and joint resolution bills as these represent the vast majority and are the forms used for all legislation concerning the public with the goal of becoming law. Our analysis is also limited to data from the current congressional term (2013-2015), i.e. the 113th congress, however some historical session data was studied to better understand the most recent data. Bills go through many stages before they become law, as shown in the table to the right.

Bill Phase	Description
INTRODUCED	Starting phase.
REFERRED	Referred to committees.
REPORTED	Come out of committee back to the house to reveal findings.
PASS_OVER	Voted on in the originating house and it passed.
PASS_BACK	Passed back to the originating house with suggestions for amendments.
SUCCESSFUL	End phase: passed in both houses and/or signed by the President.
FAILED	End phase: stopped and no further progress possible. (veto or failed vote)

Note: there is a many-to-one relationship between bill states to bill phase. For example, a bill could be passed by only the House of Representatives or only the Senate; these would both map to PASS_OVER. A full description of the fine-grained bill states is available at: <https://github.com/unitedstates/congress/wiki/bills>

There are 8750 bills in the current session, however only 200 of them have reached a completed state (whether that be successful or unsuccessful). This is because the majority of bills die due to lack of action, either because of opposition, loss of interest or changing priorities, and there is no special state to determine this. To handle this, we only included bills in our dataset where either a definitive action has been taken, e.g. enactment or failure, or no action has been made in the last four month. The final data set statistics are tabulated in the table to the right.

Number of congressmen	538
Number of bills	5041
Number of votes	775
Number of unsuccessful votes	8
Number of bills that get out of committee	15%
Of those 15%, how many get enacted	15%

Initial Approaches for Determining Bill Properties

We explored various static, inherent bill properties that could be used as indicators for success immediately after the bill of was introduced to congress. These can be categorized into the following groups.

Sponsor Properties

A polarized congress would probably mean that a bill’s sponsor and the groups they are a part of would have an impact on the bill’s outcome. Thus, we explored different dimensions that could be used to infer the influence or leadership a legislator might have. Some basic statistics about bill sponsors is tabulated in the table below.

As you can see most legislators hold at least one chair position in some committee or subcommittee, there is both a wide range in the number of bills that a legislator introduces, as well as the number of terms they have served. Since there are only a few leadership positions in congress, it make sense that the mean there is low. Most legislators are also involved in multiple committees.

	mean	median	std. dev	(min, max)
No. of chairs:	0.81	1.00	0.85	(0, 5)
No. of bills:	13.10	10.00	10.84	(0, 74)
No. of terms:	5.43	4.00	4.54	(1, 30)
No. leadership:	0.05	0.00	0.41	(0, 5)
No. of committees:	6.89	6.00	4.20	(0, 23)

We used these dimensions as well as attributes based on the congressional network to see how they affected bill outcome. These include: the number of terms a legislator has served, the number of leadership roles the legislator has held, their rank in the committees they are a member of, their PageRank score in the directed graph, where nodes are legislators and edges are made between members of committees to the chair and co-chair of that committee, and their betweenness scores in the undirected graph where edges are created between members of the same committee. This sponsor data was then plotted against the sponsor’s bill success rate.

Congressional Network Analysis

In order to determine a legislator’s PageRank and betweenness centrality score, we put together two graphs. As mentioned early, GovTrack.us calculates a legislator’s “leadership score” using the PageRank algorithm on a co-sponsorship network, i.e. a network where legislators are connected to the sponsors of the bills they have co-sponsored. However when a congressional session begins, there is no data on co-sponsorship. As our goal is to develop bill properties that are not affected by the bills themselves, we decided on the structural underpinnings of congress to calculate a “leadership score,” i.e. congressional committees and their membership. Thus, our two graphs are as follows.

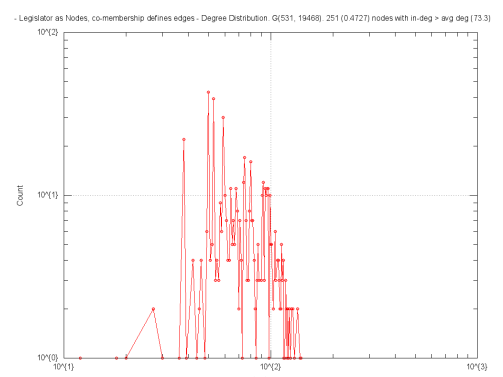
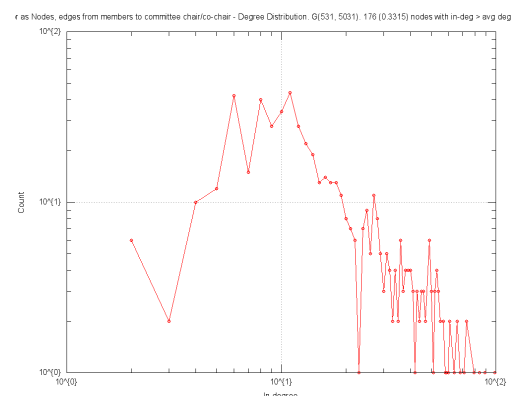
Chair-Membership Graph

This is a directed graph where nodes are legislators and members of committees have directed edges to both the chair and co-chair of their committee. This gives us a graph with 531 nodes and 5031 edges, and the highest degree node having 99 edges. We used this graph to calculate a legislator’s PageRank score. The in-degree distribution is plotted in the Figure 1a. You can see that ~20 is the highest frequency degree, which makes sense since we know a legislator holds at least one chair position on average. The remainder of the graph seems to follow somewhat of a power law degree distribution, but most nodes have some amount of influence.

Co-membership Graph

This is an undirected graph where legislators are connected if they are members of the same committee. This gives us 531 nodes and 19,468 edges, with the highest degree node having 140 edges. We used this graph to calculate a betweenness centrality score, as we thought this might give us insight into whether a legislator is a connector of committees/groups within congress. The degree distribution for this network was plotted in Figure 1b. Apart from six legislators, the remainder all are connected. The mode of the degree counts is 45 degree.

Figure 1 a, b (top, bottom)



Co-sponsor Makeup and Properties

We felt that the co-sponsor make up of a bill might also have an interesting affect on the bill's outcome. We were primarily interested in understanding how the partisanship of a bill's co-sponsors related to bill outcome. Thus we plotted the percentage of republican co-sponsors against the probability that a bill with that republican percentage succeeds. We also computed the "group rank" of the co-sponsors by averaging the ranks of all of the co-sponsors and plotted this against a bill's likelihood of success.

Sponsor and Co-sponsor Campaign Contributions

The link between campaign financing and bill success is another area we investigated. One might assume that the money that goes into a bill, in various activities like lobbying, publicizing, on interest groups, might affect a bill's chance of success. The Federal Electoral Commission provides data on all contributions made by a PAC, party committee, candidate committee, or other federal committee to particular candidates. We used this information, combined with a graph linking sponsors and their bills, to understand how it might relate to bill outcome.

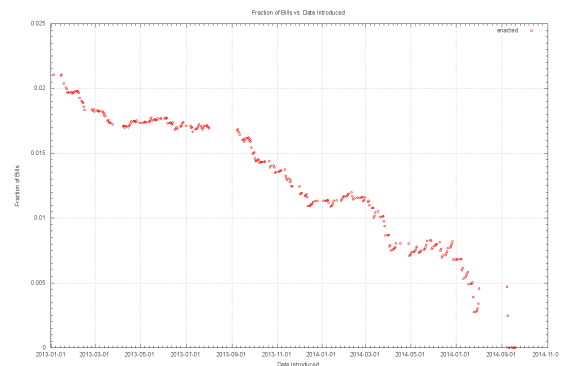
Referral Committee Properties

Another important aspect of a bill is the committee or committees it gets referred to. As explained earlier, only 15% of bills ever get reported out of committee, so this is a major area of attrition. To understand how this might affect a bill's outcome, we calculated the percentage of a bill's co-sponsors that were part of the committee that the bill was referred to. We also used 112th congressional data to calculate the committee's historic success rate and we used this as input into our prediction model.

Preliminary Findings

Introduction date.

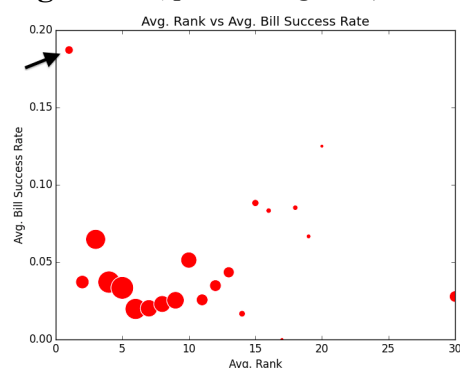
To the right is a graph that shows the fraction of successful bills out of the bills that were introduced on a given day. There is a clear trend downward as the congressional session moves forward, thus the later a legislator introduces a bill the lower the likelihood of success.



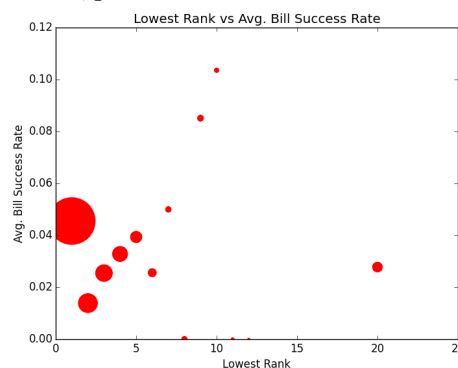
Sponsor properties v. sponsor's bill success rate

Below we plot the different attributes of the bill's sponsor against the average success rate of a sponsor with that particular attribute value. The size of the circle corresponds to the number of sponsors that have the specified attribute. You can see that much of the data does not show any definitive correlation, however there are some key findings worth noting. In Figure 2a, you can see that when a sponsor's average ranking is a 1, meaning they are a chair of all of the committees they are a part of, the chances of success double. In Figure 2c, its possible that there is a correspondence between a sponsor's length of time in office and their success rate, as legislators who have spent between 12 and 18 terms in office have almost double the success rate of their younger counterparts. There are few data points however, so it is hard to draw any real conclusions.

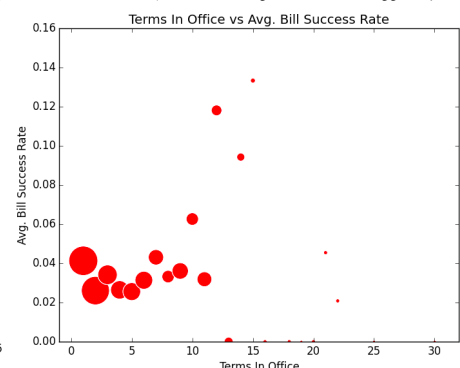
Figure 2 a (sponsor's avg. rank)

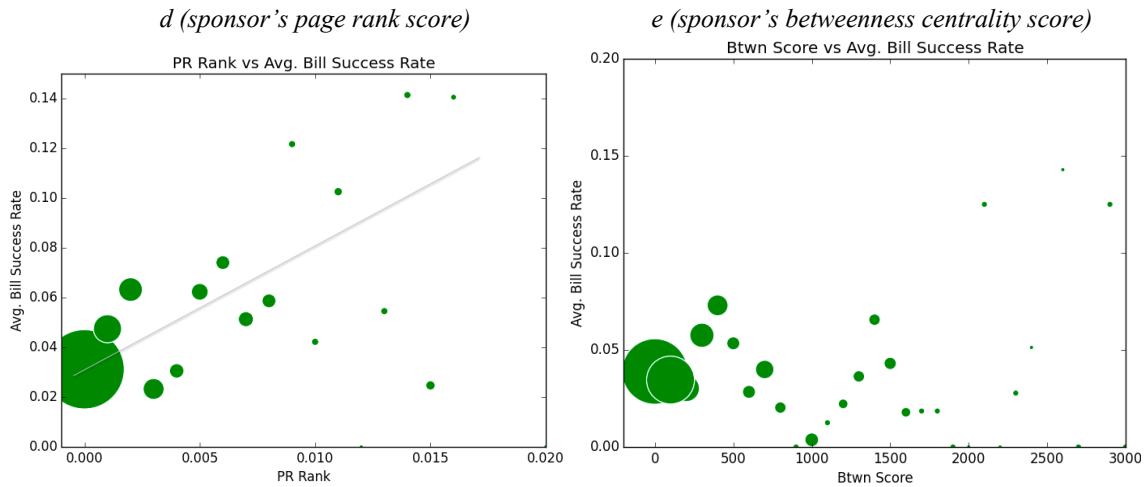


b (sponsor's lowest rank in committee)



c (number of terms in office)





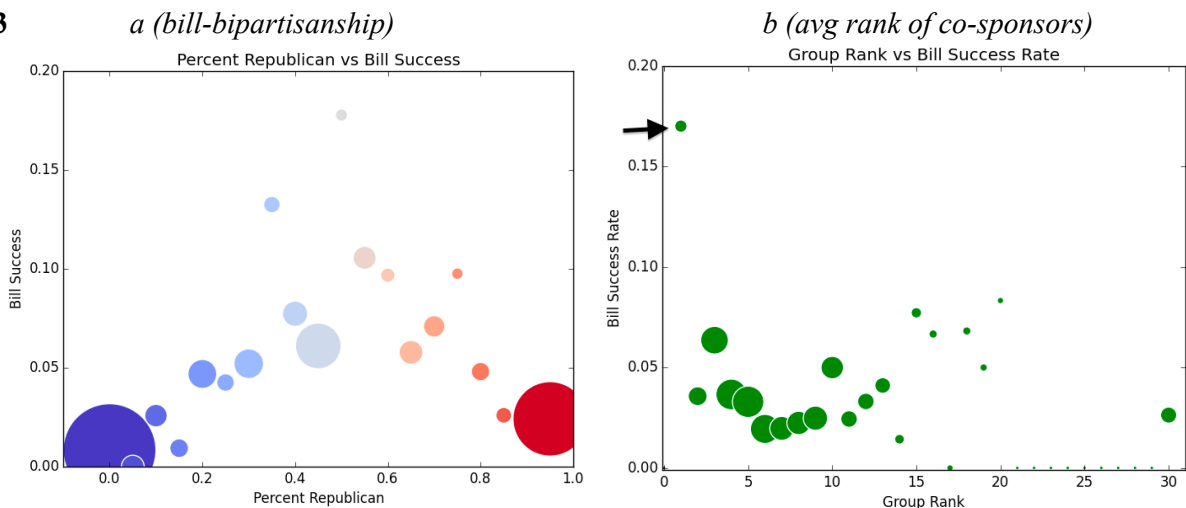
In Figure 2d we have the pagerank and the corresponding bill success rate. Again, it seems that there could be a correspondence, but the data is limited. Nonetheless, we can use this information about a bill's sponsor as input to our prediction model.

Co-sponsor properties v. bill success rate

In Figure 3a, below, you can see our results for plotting the percentage of Republican co-sponsors versus probability of bill success. We can see that the majority of bills are either mostly Democrat or mostly Republican, and that the all Democrat co-sponsored bill has the lowest chance of succeeding. This is not a surprising result with a polarized congress with a Republican majority. You can also see that there is an increase in the likelihood of success when republicans and democrats work together, fairly evenly on a bill.

In Figure 2b, we plotted the average ranking of the co-sponsors with bill success. You can see that if all of the co-sponsors have a rank of 1 in their respective committees, there is higher chance of succeeding, however again there are very few data points, thus it is unclear whether this is a pattern.

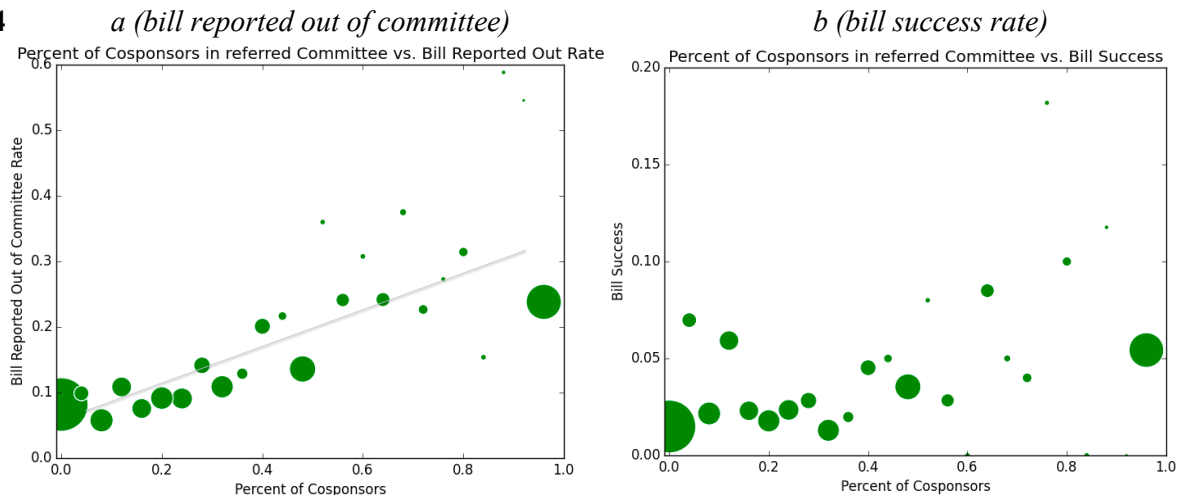
Figure 3



Referral committee properties v. bill outcome

In Figure 4a and b, you can see the results of mapping the percentage of co-sponsors that are members of the referral committee to both the rate at which the bill is reported out of committee and the overall success (enactment) rate. While there could be a potential trend in Figure 4b, there definitely seems to be a trend in Figure 4a. This would make sense as the bill's co-sponsors have more control over this step in the process when a large percentage of them are in the referral committee.

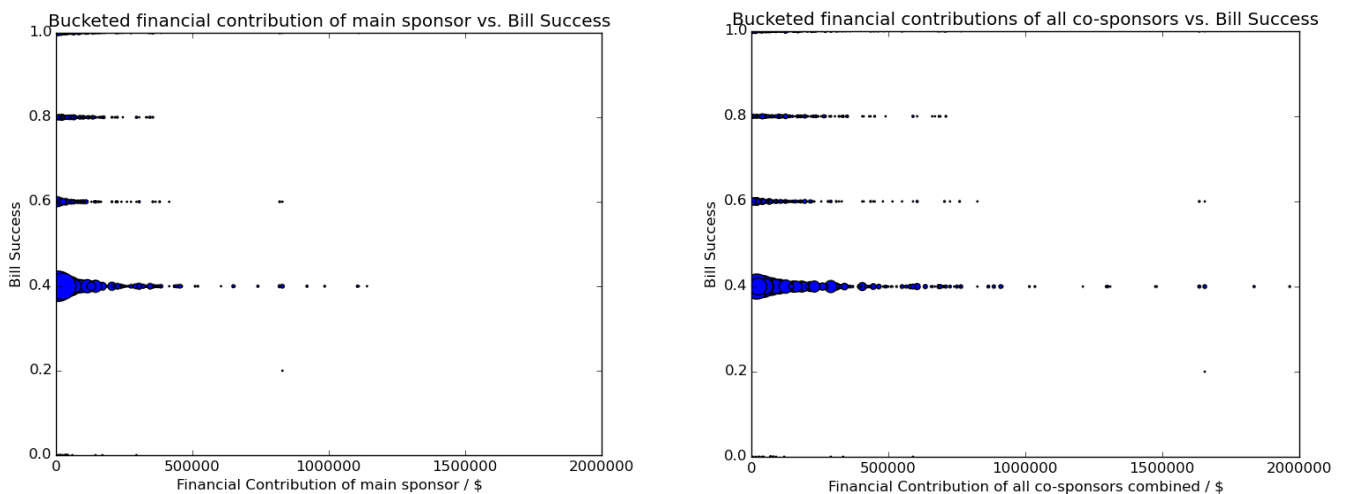
Figure 4



Spending / Financial Contributions

The plots below show the amount that the bill's main sponsor and co-sponsors collectively received against the bill's success. To reduce the clutter in the data set, the amounts were bucketed into intervals of \$5000 and this was plotted against a bill success score. The \$5000 interval was chosen because it was large enough that each bucket had more than just a handful of bills and not so large that all bills went into only a few buckets. The size of each circle is proportional to the number of bills that fell into that bucket.

Most bills have not reached a completed state and the resulting plot has nearly all the data points on the bottom axis. Instead, bill success was bucketed according to the main stages, seen in the Data section.



There was little correlation between the amount of financial contributions sponsors received and the success of their bills. Most bills received \$20,000 from their main sponsor and \$50,000 from all their co-sponsors. There were no changes in the distribution of success for different spending amounts. Surprisingly, bills with a large amount of funding, say over \$500,000 were no more successful than those with \$20,000.

Through this initial investigation, we were able to gain some interesting findings that help us understand how different bill properties as well as a bill's context within the congressional network can affect its outcome. Most notably, the introduction date, the bipartisanship of the co-sponsors, and the PageRank score of the bill's sponsor seem to have some correlation to a bill's success. We use this information to help inform our final prediction model.

Dynamic Features and Matrix Factorization for Predicting Individual Votes

Additional areas we investigated take a step away from strictly examining properties of the bills themselves, and incorporate more dynamic information about Congress members and their behaviors in our prediction efforts. A

single bill can pass through multiple rounds of voting; it has to pass through both Houses and may require additional rounds if it fails in either House or due to amendments. Analysing the dataset showed that over 98% of voting rounds result in a bill passing, which is remarkably high.

One dynamic feature we modeled is the voting behavior of individual legislators. To do this we used matrix factorization, a technique used by recommender systems to predict users' ratings of items based on a sparse set of known ratings [4]. The basic idea is that a $|U| \times |I|$ user-item rating matrix R can be factored into matrices P and Q , where P is a $|U| \times k$ matrix representing a set of k latent features representing the user's preferences, and Q is a $|I| \times k$ matrix representing each item by the same set of k latent features. The product $P * Q^T = \hat{R}$ is used to predict each user's rating across all items. P and Q can be found by using gradient descent to minimize the mean squared error of the resulting rating predictions.

To perform our analysis, a simple matrix factorization method with random initialization (with values between 0 and 1) of a set of $k=40$ latent features and subsequent gradient descent was used. Data from the current and previous two sessions of congress was used. The algorithm was run until the improvement of each iteration became sufficiently small (less than a 0.1% decrease in mean squared error). The model was also improved by accounting for bias of each legislator as suggested in [4], in which the average deviation from the mean score for each bill is added to the legislator's predicted score, and static features known about each legislator (gender and party affiliation) using the techniques presented in [5], in which static features are added to the P matrix that remain unchanged by gradient descent.

In order to produce features for the machine learned predictor, rating matrices were created in which all votes for the bill in question were hidden except the votes of the sponsor and cosponsors. The algorithm then predicted rating scores for the remaining congress members for that bill based on the votes on other bills still present in the rating matrix. Features provided to the machine-learned models include the average predicted rating, standard deviation of predicted ratings, and a histogram of ratings using values at the 10th, 25th, 50th, 75th, and 90th percentiles.

Making predictions via Machine Learning

The final phase of our analysis was to input the various properties discussed earlier into a machine learning algorithm in order to make predictions on a bill's success, and then use feature selection to identify the best features. These results were compared against our own understanding of which properties were most useful.

Three machine learning algorithms were used so as to ascertain a better picture than just using one. Both tasks were binary classification problems: predict whether a bill will come out of committee or not and predict whether a bill, once out of committee, will succeed or not. Logistic Regression, Support Vector Machines and Naive Bayes were chosen. In addition to the accuracy, for a more rounded measure of the performance of each algorithm, the precision, recall and F1 score were also calculated. Forward search was implemented to discover which were the most informative features.

The following properties of a bill were used as features: the number of cosponsors, the month it was introduced, the number of rounds of voting the bill has seen, the number of successful rounds of voting the bill has had, amount of financial contribution, bipartisan score, the percentage of cosponsors that are members of the committee it is referred to, historical rate for success, historical rate for getting out committee and statistics from matrix factorization: mean, standard deviation, histogram values at 10%, 25% 50%, 75% and 90%.

Matrix Factorization and Machine Learning Results

Predicting individual votes

Matrix factorization proved to be very successful in predicting votes when given a sufficient number of known votes for each bill. When examining votes from only the current session of congress, it became apparent that most legislators voted consistently for or against all bills presented. Thus a baseline method of predicting that a legislator

will always vote as they most frequently do on each bill had surprisingly high success, which matrix factorization was barely able to beat. However, when using data from the past 3 sessions of congress, matrix factorization performed significantly better; the variation in each congress member’s votes increased, although a fair number of legislators voted consistently enough one way or the other that the method had reasonable performance. The tables below show performance of both methods on the current session of congress and on the past three sessions; it is clear that matrix factorization performed better or at about the same level as the baseline method in terms of accuracy, precision, and recall. Figures 6 and 7 show accuracy of predictions computed on a per-legislator basis, and show that matrix factorization shows much better accuracy in predicting votes for the majority of legislators. In measuring performance, 15% of the data was withheld as a test set and the methods were trained on the remaining 85%.

	<i>Performance on current session of Congress</i>			<i>Performance on past 3 sessions of Congress</i>		
	Accuracy	Precision	Recall	Accuracy	Precision	Recall
M.F.	0.95	0.96	0.96	0.95	0.96	0.96
Baseline	0.93	0.97	0.91	0.77	0.83	0.78

However, as the number of known votes for a bill is decreased, the performance of matrix factorization also decreases significantly; Figure 5 shows the decrease in performance as the number of (randomly chosen) votes withheld per bill increases. When only the votes of sponsors and cosponsors were given, accuracy decreased to 0.47, precision to 0.57, and recall to 0.50. This is because only votes in favor of a bill are provided, and also because most bills have a fairly small number of cosponsors (mean: 16.6, median: 6). Thus, as a bill progresses through the legislative process and the votes of more legislators are known, matrix factorization is able to better predict any unknown votes.

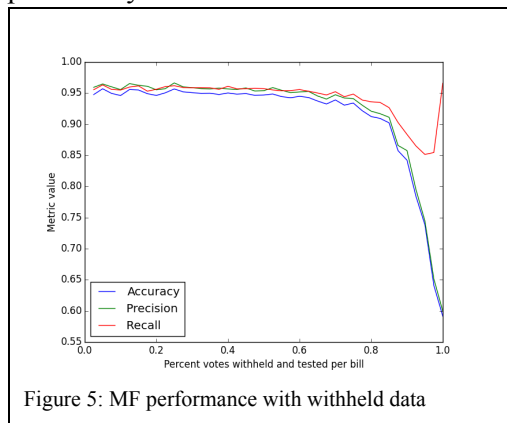


Figure 5: MF performance with withheld data

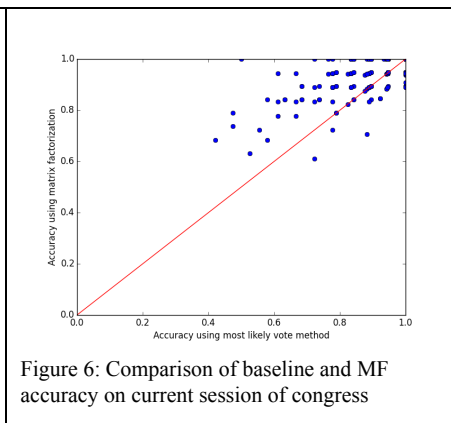


Figure 6: Comparison of baseline and MF accuracy on current session of congress

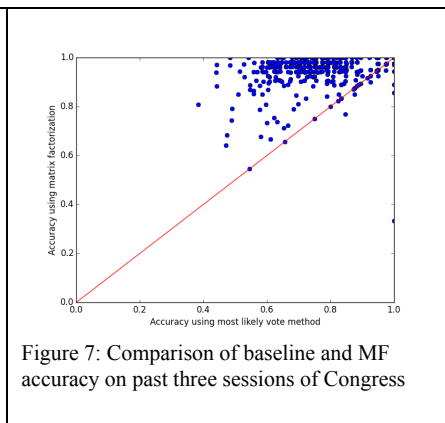


Figure 7: Comparison of baseline and MF accuracy on past three sessions of Congress

The performance of matrix factorization is limited by several factors. When using only votes from sponsors and cosponsors, there are no known negative ratings of the bills; thus there is an overall positive bias in the predicted scores. In addition, there may be a selective bias within congress itself for which bills are chosen to be voted upon; some bills that are not enacted are never even voted upon, causing the proportion of successful bills in the set of bills that underwent vote to be higher than the proportion of successful bills across all proposed bills. In many cases, legislators may withhold their vote instead of explicitly voting against a bill, further biasing the available data; 60.0% of observed votes are cast in favor of the proposed bill.

Machine Learning Predictions

The results of the machine learning predictions were mixed. Precision, recall and F1 scores were in-line with accuracy scores. Naive Bayes performed very well when all the features were used with a 5% and 20% improvement over the baseline for the two different predictions. SVM did no better than the baseline, which is to always predict ‘No’. Logistic regression fared poorly. The bipartisan score of a bill was found to be an important feature and this followed a binomial distribution. Logistic regression can fit a decision boundary to linear

correlations but struggles with other distributions. Logistic Regression’s low recall score of 21% indicates that finds it hard to identify successful bills.

Accuracy	<i>Will a bill will get out of committee?</i>				<i>Will a bill, once out of committee, get enacted?</i>			
	Logistic Regression	SVM	Naive Bayes	Baseline	Logistic Regression	SVM	Naive Bayes	Baseline
All features	21.06%	84.08%	88.01%	84.17%	79.72%	82.17%	98.60%	82.17%
Excluding voting rounds	21.06%	83.94%	70.66%		79.72%	82.17%	81.12%	

The reason Naive Bayes was chosen is that the dataset has a heavy bias towards classifying ‘No’ as 85% of the bills did not get enacted. Naive Bayes, with Laplace smoothing, is able to counteract this by using a generative model.

Running with all features	<i>Will a bill will get out of committee?</i>			<i>Will a bill, once out of committee, get enacted?</i>		
	Logistic Regression	SVM	Naive Bayes	Logistic Regression	SVM	Naive Bayes
Precision	72%	71%	88%	71%	68%	99%
Recall	21%	84%	88%	80%	82%	99%
F1 score	16%	77%	86%	74%	74%	99%

Running forward search for feature selection determined that two features were most important in the predictions: 'the number of rounds of voting the bill has seen' and 'the number of successful rounds of voting the bill has had' implies that the bill has come out of committee. This makes sense as bills can only be voted on once they've been reported out of committee. A non-zero value in either of these implies the bill must be out of committee. In other words, bills are only sent for a vote if there is a very high chance that it will succeed. There is overhead in organising a vote as well as the risk of negative publicity if a vote fails, which dissuades congressmen from putting a bill to vote until they are sure it will pass.

These two features are non-static; they change as the bill progresses through stages. The predictions were repeating after excluding them and having a feature set of only static properties. In this set of predictions, there was no change to the results of Logistic Regression but Naive Bayes performed much worse. SVM was relatively unchanged. The bipartisan score and matrix factorization mean score, and percent of cosponsors in the committee a bill is referred to were the next most important features. The amount of financial contribution a bill received was one of the worst performing features. This is not surprising given earlier results from plotting it against bill success.

Conclusion and Further Work

After investigating a wide number of characteristics and using multiple techniques to predict the success of Congressional bills, we have had moderate success at predicting how successful bills will be and how Congress members will vote on them. As mentioned previously, we found that the most helpful features in predicting a bill’s success are dynamic features that change as a bill makes progress towards passage. This intuitively makes sense; as we know more information about a bill, we are able to better predict the outcome.

This is similar to the trend seen in predicting legislators’ individual votes; knowing how more Congress members vote results in improved prediction of how remaining members will vote. As mentioned, matrix factorization techniques are able to successfully predict legislator’s votes, but performance degrades quickly when only a small number of legislators’ opinions are known beforehand.

Several of the bill characteristics we investigated did not show a clear correlation with bill success. However, there did seem to be some relationships with the bipartisanship of cosponsors, sponsor PageRank, and bill introduction date and the final bill outcome. Using historical congressional data would help show whether these are actual,

strong correlations, and would also help further train our prediction model. Another factor that may have contributed to the uncertain correlations, is the nature of what kinds of bills get passed. In further work it would be helpful to separate enacted bills into those that pass easily, those that were controversial, and those that were never going to make it. We would then be able to clarify when the bill properties would become important.

In the case of spending and campaign contributions, we did not observe a clear correlation for several possible reasons. First, not all contributions were required to be reported. Second, funding may not actually in itself be a good criterion of success. More funding may help, but given that the legislators have made it to Washington, they must already be very good at what they do and so are less affected by financial contributions. Third, given the amount of cynicism around government processes, there may be other avenues that this project was unaware of and did not consider.

More recently, the FEC published data on any campaign committee, party committee or leadership PAC that receives more than \$16,000 bundled by a lobbyist. This data is at committee level. There is no breakdown into how each committee utilised the contribution and what proportion was directed at a congressmen. These disclosures are relatively new and have not filtered through into the candidates data file that we used.

Further investigation could include training the model multiple times with different features using cross validation and for datasets from different congress sessions. Matrix factorization methods could continue to be improved, perhaps using additional static features of both bills and legislators. In addition, more types of features could be explored, such as information about the content and topic of each bill. Additional improvements could include using more data (e.g. more historical data from Congress sessions), attempting to predict a probability of bill success rather than a binary prediction, and identifying bills that always succeed (such as post office designations or other uncontroversial bills).

In conclusion, we identified a number of features that appear to correlate with bill success and were successfully used by a machine-learned model to predict the success or failure of a bill.

REFERENCES

- (1) Bill Prognosis Analysis. *GovTrack.us*. 2013
- (2) Jure Leskovec, Daniel Huttenlocher, Jon Kleinberg. "Predicting positive and negative links in online social networks." Proceedings of the 19th international conference on World wide web, April 26-30, 2010, Raleigh, North Carolina, USA
- (3) R. Guha, Ravi Kumar, Prabhakar Raghavan, Andrew Tomkins. "Propagation of trust and distrust." Proceedings of the 13th international conference on World Wide Web, May 17-20, 2004, New York, NY, USA
- (4) Koren, Yehuda, Robert Bell, and Chris Volinsky. "Matrix factorization techniques for recommender systems." *Computer* 42.8 (2009): 30-37.
- (5) G. Takács et al., "Major Components of the Gravity Recommendation System," *SIGKDD Explorations*, vol. 9, 2007, pp. 80-84.
- (6) Paterek, Arkadiusz. "Improving regularized singular value decomposition for collaborative filtering." *Proceedings of KDD cup and workshop*. Vol. 2007. 2007.
- (7) Terkel, Amanda. "112th Congress Set To Become Most Unproductive Since 1940s." The Huffington Post. TheHuffingtonPost.com, 28 Dec. 2012. Web. 16 Oct. 2014.
- (8) "How Congress Became the Most Polarized and Unproductive It's Ever Been." Washington Post. The Washington Post. Web. 16 Oct. 2014.
- (9) Tauberer, Joshua. "Observing the Unobservables in the U.S. Congress", presented at Law Via the Internet 2012, Cornell Law School, October 2012.